

Multiple Testing for Neuroimaging via Hidden Markov Random Field

Hai Shu,¹ Bin Nan,^{1,*} and Robert Koeppe²

¹Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

²Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

*email: bnan@umich.edu

SUMMARY. Traditional voxel-level multiple testing procedures in neuroimaging, mostly p -value based, often ignore the spatial correlations among neighboring voxels and thus suffer from substantial loss of power. We extend the local-significance-index based procedure originally developed for the hidden Markov chain models, which aims to minimize the false nondiscovery rate subject to a constraint on the false discovery rate, to three-dimensional neuroimaging data using a hidden Markov random field model. A generalized expectation–maximization algorithm for maximizing the penalized likelihood is proposed for estimating the model parameters. Extensive simulations show that the proposed approach is more powerful than conventional false discovery rate procedures. We apply the method to the comparison between mild cognitive impairment, a disease status with increased risk of developing Alzheimer’s or another dementia, and normal controls in the FDG-PET imaging study of the Alzheimer’s Disease Neuroimaging Initiative.

KEY WORDS: Alzheimer’s disease; False discovery rate; Generalized expectation–maximization algorithm; Ising model; Local significance index; Penalized likelihood.

1. Introduction

In a seminal article, Benjamini and Hochberg (1995) introduced false discovery rate (FDR) as an alternative measure of Type I error in multiple testing problems to the family-wise error rate (FWER). They showed that the FDR is equivalent to the FWER if all null hypotheses are true and is smaller otherwise, thus FDR controlling procedures potentially have a gain in power over FWER controlling procedures. FDR is defined as the expected proportion of false rejections among all rejections. The false nondiscovery rate (FNR; Genovese and Wasserman, 2002), the expected proportion of falsely accepted hypotheses, is the corresponding measure of Type II error. The traditional FDR procedures (Benjamini and Hochberg, 1995, 2000; Genovese and Wasserman, 2004), which are p -value based, are theoretically developed under the assumption that the test statistics are independent. Although these approaches are shown to be valid in controlling FDR under certain dependence assumptions (Benjamini and Yekutieli, 2001; Farcomeni, 2007; Wu, 2008), they may suffer from severe loss of power when the dependence structure is ignored (Sun and Cai, 2009). By modeling the dependence structure using a hidden Markov chain (HMC), Sun and Cai (2009) proposed an oracle FDR procedure built on a new test statistic, the local index of significance (LIS), and the corresponding asymptotic data-driven procedure, which are optimal in the sense that they minimize the marginal FNR subject to a constraint on the marginal FDR. Following the work of Sun and Cai (2009), Wei et al. (2009) developed a pooled LIS (PLIS) procedure for multiple-group analysis where different groups have different HMC dependence structures, and proved the optimality of the PLIS procedure. Either the LIS procedure

or the PLIS procedure only handles the one-dimensional dependency. However, problems with higher dimensional dependence are of particular practical interest in analyzing imaging data.

FDR procedures have been widely used in analyzing neuroimaging data, such as positron emission tomography (PET) imaging and functional magnetic resonance imaging (fMRI) data (Genovese, Lazar, and Nichols, 2002; Chumbley and Friston, 2009; Chumbley et al., 2010, among many others). We extend the work of Sun and Cai (2009) in this article by developing an optimal LIS-based FDR procedure for three-dimensional (3D) imaging data using a hidden Markov random field model (HMRF) for the spatial dependency among multiple tests. Existing methods for correlated imaging data, for example, Zhang, Fan, and Yu (2011) are not shown to be optimal, i.e., minimizing FNR.

HMRF model is a generalization of HMC model, which replaces the underlying Markov chain by Markov random field. A well-known classical Markov random field with two states is the Ising model. In particular, the two-parameter Ising model, whose formal definition is given in equation (1), reduces to the two-state Markov chain in one-dimension (Bremaud, 1999). The Ising model and its generalization with more than two states, the Potts model, have been widely used to capture the spatial structure in image analysis; see Bremaud (1999), Winkler (2003), Zhang et al. (2008), Huang et al. (2013) and Johnson et al. (2013), among others. In this article, we consider a hidden Ising model for each area based on the Brodmann’s partition of the cerebral cortex (Garey, 2006) and subcortical regions of the human brain, which provides a natural way of modeling spatial correlations for neuroimaging

data. To the best of our knowledge, this is the first work that introduces the HMRF-LIS based FDR procedure to the field of neuroimaging.

We propose a generalized expectation-maximization algorithm (GEM; Dempster et al., 1977) to search for penalized maximum likelihood estimators (Ridolfi, 1997; Ciuperca, Ridolfi, and Idier, 2003; Chen, Tan, and Zhang, 2008) of the hidden Ising model parameters. The penalized likelihood prevents the unboundedness of the likelihood function, and the proposed GEM uses Monte Carlo averages via Gibbs sampler (Geman and Geman, 1984; Roberts and Smith, 1994) to overcome the intractability of computing the normalizing constant in the underlying Ising model. Then the LIS-based FDR procedures can be conducted by plugging in the estimates of the hidden Ising model parameters. In what follows, we use the term “HMRF” to refer to the 3D hidden Ising model.

The article is organized as follows. In Section 2, we introduce the HMRF model, i.e., the hidden Ising model, for 3D imaging data. We provide the GEM algorithm for the HMRF parameter estimation and the implementation of the HMRF-LIS-based data-driven procedures in Section 3. In Section 4, we conduct extensive simulations to compare the LIS-based procedures with conventional FDR methods. In Section 5, we apply the PLIS procedure to the ^{18}F -Fluorodeoxyglucose PET (FDG-PET) image data of the Alzheimer’s Disease Neuroimaging Initiative (ADNI), which finds more signals than conventional methods.

2. A Hidden Markov Random Field Model

Let S be a finite lattice of N voxels in an image grid, usually in a 3D space. Let $\Theta = \{\Theta_s \in \{0, 1\} : s \in S\}$ denote the set of latent states on S , where $\Theta_s = 1$ if the null hypothesis at voxel s is false and $\Theta_s = 0$ otherwise. For simplicity, we follow Sun and Cai (2009) to call hypothesis s to be nonnull if $\Theta_s = 1$ and null otherwise. We also call voxel s to be a signal if $\Theta_s = 1$ and noise otherwise. Let Θ be generated from a two-parameter Ising model with the following probability distribution

$$P_{\Phi}(\theta) = \frac{1}{Z(\Phi)} \exp\{\Phi^T \mathbf{H}(\theta)\} \\ = \frac{1}{Z(\beta, h)} \exp\left\{\beta \sum_{(s,t)} \theta_s \theta_t + h \sum_{s \in S} \theta_s\right\}, \quad (1)$$

where $Z(\Phi)$ is the normalizing constant, $\Phi = (\beta, h)^T$, $\mathbf{H}(\theta) = (\sum_{(s,t)} \theta_s \theta_t, \sum_{s \in S} \theta_s)^T$, and (s, t) denotes all the unordered pairs in S such that for any s, t is among the six nearest neighbors of voxel s in a 3D setting. This model possesses the Markov property:

$$P_{\Phi}(\theta_s | \theta_{S \setminus \{s\}}) = P_{\Phi}(\theta_s | \theta_{\mathcal{N}(s)}) \\ = \frac{\exp\{\theta_s (\beta \sum_{t \in \mathcal{N}(s)} \theta_t + h)\}}{1 + \exp\{\beta \sum_{t \in \mathcal{N}(s)} \theta_t + h\}},$$

where $S \setminus \{s\}$ denotes the set S after removing s , and $\mathcal{N}(s) \subset S$ is the nearest neighborhood of s in S . Some parameter interpretations of β and h are given in Web Appendix A.

We assume the observed z -values $\mathbf{X} = \{X_s : s \in S\}$ are independent given $\Theta = \theta$ with

$$P_{\Phi}(\mathbf{x} | \theta) = \prod_{s \in S} P_{\Phi}(x_s | \theta_s), \quad (2)$$

where $P_{\Phi}(x_s | \theta_s)$ denotes the following distribution

$$X_s | \Theta_s \sim (1 - \Theta_s) N(\mu_0, \sigma_0^2) + \Theta_s \sum_{l=1}^L p_l N(\mu_l, \sigma_l^2) \quad (3)$$

with $(\mu_0, \sigma_0^2) = (0, 1)$, unknown parameters $\Phi = (\mu_1, \sigma_1^2, p_1, \dots, \mu_L, \sigma_L^2, p_L)^T$, $\sum_{l=1}^L p_l = 1$ and $p_l \geq 0$. In particular, the z -value X_s follows the standard normal distribution under the null, and the nonnull distribution is set to be the normal mixture that can be used to approximate a large collection of distributions (Magder and Zeger, 1996; Efron, 2004). The number of components L in the nonnull distribution may be selected by, for example, the Akaike or Bayesian information criterion. Following the recommendation of Sun and Cai (2009), we use $L = 2$ for the ADNI image analysis.

Markov random fields (MRFs; Bremaud, 1999) are a natural generalization of Markov chains (MCs), where the time index of MC is replaced by the space index of MRF. It is well known that any one-dimensional MC is an MRF, and any one-dimensional stationary finite-valued MRF is an MC (Chandgotia et al., 2014). When S is taken to be one-dimensional, the above approach based on (1)–(3) reduces to the HMC method of Sun and Cai (2009).

3. Hidden Markov Random Field LIS-Based FDR Procedures

Sun and Cai (2009) developed a compound decision theoretic framework for multiple testing under HMC dependence and proposed LIS-based oracle and data-driven testing procedures that aim to minimize the FNR subject to a constraint on FDR. We extend these procedures under HMRF for image data. The oracle LIS for hypothesis s is defined as $LIS_s(\mathbf{x}) = P_{\Phi}(\Theta_s = 0 | \mathbf{x})$ for a given parameter vector Φ . In our model, $\Phi = (\Phi^T, \Phi^T)^T$. Let $LIS_{(1)}(\mathbf{x}), \dots, LIS_{(N)}(\mathbf{x})$ be the ordered LIS values and $\mathcal{H}_{(1)}, \dots, \mathcal{H}_{(N)}$ the corresponding null hypotheses. The oracle procedure operates as follows: for a prespecified FDR level α ,

$$\text{let } k = \max \left\{ i : \frac{1}{i} \sum_{j=1}^i LIS_{(j)}(\mathbf{x}) \leq \alpha \right\}, \\ \text{then reject all } \mathcal{H}_{(i)}, i = 1, \dots, k. \quad (4)$$

Parameter Φ is unknown in practice. We can use the data-driven procedure that simply replaces $LIS_{(i)}(\mathbf{x})$ in (4) with $\widehat{LIS}_{(i)}(\mathbf{x}) = P_{\hat{\Phi}}(\Theta_{(i)} = 0 | \mathbf{x})$, where $\hat{\Phi}$ is an estimate of Φ .

If all the tests are partitioned into multiple groups and each group follows its own HMRF, in contrast to the separated LIS (SLIS) procedure that conducts the LIS-based FDR procedure separately for each group at the same FDR level α and then combines the testing results, we follow Wei et al. (2009) to propose a pooled LIS (PLIS) procedure that is more efficient

in reducing the global FNR. The PLIS follows the same procedure as (4), but with $LIS_{(1)}, \dots, LIS_{(N)}$ being the ordered test statistics from all groups.

Note that the model homogeneity, which is required in Sun and Cai (2009) and Wei et al. (2009) for HMCs, fails to hold for the HMRF model. In other words, $P(\Theta_s = 1)$ for the interior voxels with six nearest neighbors are different to those for the boundary voxels with less than six nearest neighbors. We show the validity and optimality of the oracle HMRF-LIS-based procedures in Web Appendix B.

We now provide details of the LIS-based data-driven procedure for 3D image data, where the parameters of the HMRF model need to be estimated from observed test data.

3.1. A Generalized EM Algorithm

The observed likelihood function under HMRF, $L(\Phi|\mathbf{x}) = P_\Phi(\mathbf{x}) = \sum_{\Theta} P_\Phi(\mathbf{x}|\Theta)P_\Theta(\Theta)$, is unbounded (see Web Appendix C for details). One solution to avoid the unboundedness is to replace the likelihood by a penalized likelihood (Ridolfi, 1997; Ciuperca et al., 2003)

$$pL(\Phi|\mathbf{x}) = L(\Phi|\mathbf{x}) \prod_{l=1}^L g(\sigma_l^2), \quad (5)$$

where $g(\sigma_l^2)$, $l = 1, \dots, L$, are penalty functions that ensure the boundedness of $pL(\Phi|\mathbf{x})$. We follow Ridolfi (1997) and Ciuperca et al. (2003) to choose

$$g(\sigma_l^2) \propto \frac{1}{\sigma_l^{2b}} \exp \left\{ -\frac{a}{\sigma_l^2} \right\}, \quad a > 0, b \geq 0,$$

where $x \propto y$ means that $x = cy$ with a positive constant c independent of any parameter. Note that (5) reduces to the unpenalized likelihood function when $a = b = 0$. When $a > 0$ and $b > 1$, the penalized likelihood approach is equivalent to setting $g(\sigma_l^2)$ to be the inverse gamma distribution, which is a classical prior distribution for the variance of a normal distribution in Bayesian statistics (Hoff, 2009). We do not impose any prior distribution here. The choice of a and b does not impact the strong consistency of the penalized maximum likelihood estimator (PMLE) based on the same penalty function for a finite mixture of normal distributions (Ciuperca et al., 2003; Chen et al., 2008). Such a penalty performs well in the simulations, though formal proof of the consistency of PMLE for hidden Ising model remains an open question.

We develop an EM algorithm based on the penalized likelihood (5) for the estimation of parameters in the HMRF model characterized by (1)–(3). We introduce unobservable categorical variables $\mathbf{K} = \{K_s : s \in S\}$, where $K_s = 0$ if $\Theta_s = 0$, and $K_s \in \{1, \dots, L\}$ if $\Theta_s = 1$. Hence, $P(K_s = 0 | \Theta_s = 0) = 1$ and we denote $P(K_s = l | \Theta_s = 1) = p_l$. From (3), we let $X_s | K_s \sim N(\mu_{K_s}, \sigma_{K_s}^2)$. To estimate the HMRF parameters $\Phi = (\phi^T, \varphi^T)^T$, $(\Theta, \mathbf{K}, \mathbf{X})$ are used as the complete data variables to construct the auxiliary function in the $(t+1)$ st iteration of EM algorithm given the observed data \mathbf{x} and the current estimated parameters $\Phi^{(t)}$:

$$Q(\Phi|\Phi^{(t)}) = E_{\Phi^{(t)}}[\log P_\Phi(\Theta, \mathbf{K}, \mathbf{X})|\mathbf{x}] + \sum_{l=1}^L \log g(\sigma_l^2),$$

where $P_\Phi(\Theta, \mathbf{K}, \mathbf{X}) = P_\phi(\Theta)P_\phi(\mathbf{X}, \mathbf{K}|\Theta) = P_\phi(\Theta) \prod_{s \in S} P_\phi(X_s, K_s | \Theta_s)$. The Q -function can be further written as follows

$$Q(\Phi|\Phi^{(t)}) = Q_1(\phi|\Phi^{(t)}) + Q_2(\varphi|\Phi^{(t)}),$$

where

$$Q_1(\phi|\Phi^{(t)}) = \sum_{\Theta} \sum_{\mathbf{K}} P_{\Phi^{(t)}}(\Theta, \mathbf{K}|\mathbf{x}) \log P_\phi(\mathbf{x}, \mathbf{K}|\Theta) + \sum_{l=1}^L \log g(\sigma_l^2)$$

and

$$Q_2(\varphi|\Phi^{(t)}) = \sum_{\Theta} P_{\Phi^{(t)}}(\Theta|\mathbf{x}) \log P_\varphi(\Theta).$$

Therefore, we can maximize $Q(\Phi|\Phi^{(t)})$ for Φ by maximizing $Q_1(\phi|\Phi^{(t)})$ for ϕ and $Q_2(\varphi|\Phi^{(t)})$ for φ , separately.

Maximizing $Q_1(\phi|\Phi^{(t)})$ under the constraint $\sum_{l=1}^L p_l = 1$ by the method of Lagrange multipliers yields,

$$p_l^{(t+1)} = \frac{\sum_{s \in S} w_s^{(t)}(l)}{\sum_{s \in S} \gamma_s^{(t)}(1)}, \quad (6)$$

$$\mu_l^{(t+1)} = \frac{\sum_{s \in S} w_s^{(t)}(l)x_s}{\sum_{s \in S} w_s^{(t)}(l)}, \quad (7)$$

$$(\sigma_l^2)^{(t+1)} = \frac{2a + \sum_{s \in S} w_s^{(t)}(l)(x_s - \mu_l^{(t+1)})^2}{2b + \sum_{s \in S} w_s^{(t)}(l)}, \quad (8)$$

where

$$w_s(l) = \frac{\gamma_s(1)p_l f_l(x_s)}{f(x_s)},$$

$$\gamma_s(i) = P_\Phi(\Theta_s = i|\mathbf{x}),$$

$$f_l = N(\mu_l, \sigma_l^2), \quad \text{and} \quad f = \sum_{l=1}^L p_l f_l.$$

For $Q_2(\varphi|\Phi^{(t)})$, taking its first and second derivatives with respect to φ , we obtain

$$\begin{aligned} U^{(t+1)}(\varphi) &= \frac{\partial}{\partial \varphi} Q_2(\varphi|\Phi^{(t)}) \\ &= E_{\Phi^{(t)}}[H(\Theta)|\mathbf{x}] - E_\varphi[H(\Theta)], \\ I(\varphi) &= -\frac{\partial^2}{\partial \varphi \partial \varphi^T} Q_2(\varphi|\Phi^{(t)}) = \text{Var}_\varphi[H(\Theta)]. \end{aligned}$$

Maximizing $Q_2(\varphi|\Phi^{(t)})$ is then equivalent to solving the non-linear equation:

$$U^{(t+1)}(\varphi) = E_{\Phi^{(t)}}[H(\Theta)|\mathbf{x}] - E_\varphi[H(\Theta)] = \mathbf{0}. \quad (9)$$

It can be shown that equation (9) has a unique solution and can be solved by the Newton-Raphson (NR) method (Stoer and Bulirsch, 2002). However, a starting point that is not close enough to the solution may result in divergence of the NR method. Therefore, rather than searching for the solution of equation (9) over all $\boldsymbol{\varphi}$, we choose a $\boldsymbol{\varphi}^{(t+1)}$ that increases $Q_2(\boldsymbol{\varphi}|\boldsymbol{\Phi}^{(t)})$ over its value at $\boldsymbol{\varphi} = \boldsymbol{\varphi}^{(t)}$. Together with the maximization of $Q_1(\boldsymbol{\phi}|\boldsymbol{\Phi}^{(t)})$, the approach leads to $Q(\boldsymbol{\Phi}^{(t+1)}|\boldsymbol{\Phi}^{(t)}) \geq Q(\boldsymbol{\Phi}^{(t)}|\boldsymbol{\Phi}^{(t)})$ and thus $pL(\boldsymbol{\Phi}^{(t+1)}|\mathbf{x}) \geq pL(\boldsymbol{\Phi}^{(t)}|\mathbf{x})$, which is termed a GEM algorithm (Dempster, Laird, and Rubin, 1977). To find such a $\boldsymbol{\varphi}^{(t+1)}$ that increases the Q_2 -function, a backtracking line search algorithm (Nocedal and Wright, 2006) is applied with a set of decreasing positive values λ_m in the following

$$\boldsymbol{\varphi}^{(t+1,m)} = \boldsymbol{\varphi}^{(t)} + \lambda_m \mathbf{I}(\boldsymbol{\varphi}^{(t)})^{-1} \mathbf{U}^{(t+1)}(\boldsymbol{\varphi}^{(t)}), \quad (10)$$

where $m = 0, 1, \dots$ and $\boldsymbol{\varphi}^{(t+1)} = \boldsymbol{\varphi}^{(t+1,m)}$ which is the first one satisfying the Armijo condition (Nocedal and Wright, 2006)

$$\begin{aligned} Q_2(\boldsymbol{\varphi}^{(t+1,m)}|\boldsymbol{\Phi}^{(t)}) - Q_2(\boldsymbol{\varphi}^{(t)}|\boldsymbol{\Phi}^{(t)}) \\ \geq \alpha \lambda_m \mathbf{U}^{(t+1)}(\boldsymbol{\varphi}^{(t)})^T \mathbf{I}(\boldsymbol{\varphi}^{(t)})^{-1} \mathbf{U}^{(t+1)}(\boldsymbol{\varphi}^{(t)}). \end{aligned} \quad (11)$$

Since $\mathbf{I}(\boldsymbol{\varphi}^{(t)})$ is positive-definite, the Armijo condition guarantees the increase of Q_2 -function. In practice, α is chosen to be quite small. We adopt $\alpha = 10^{-4}$, which is recommended by Nocedal and Wright (2006), and halve the Newton-Raphson step length each time by using $\lambda_m = 2^{-m}$.

In the GEM algorithm, Monte Carlo averages are used via Gibbs sampler to approximate the quantities of interest that are involved with the intractable normalizing constant of the Ising model. By the ergodic theorem of the Gibbs sampler (Roberts and Smith, 1994) (see Web Appendix D for details),

$$\begin{aligned} \mathbf{U}^{(t+1)}(\boldsymbol{\varphi}) &\approx \frac{1}{n} \sum_{i=1}^n (\mathbf{H}(\boldsymbol{\theta}^{(t,i,\mathbf{x})}) - \mathbf{H}(\boldsymbol{\theta}^{(i,\boldsymbol{\varphi})})), \\ \mathbf{I}(\boldsymbol{\varphi}) &\approx \frac{1}{n-1} \sum_{i=1}^n \left(\mathbf{H}(\boldsymbol{\theta}^{(i,\boldsymbol{\varphi})}) - \frac{1}{n} \sum_{j=1}^n \mathbf{H}(\boldsymbol{\theta}^{(j,\boldsymbol{\varphi})}) \right)^{\otimes 2}, \end{aligned}$$

where $\{\boldsymbol{\theta}^{(t,1,\mathbf{x})}, \dots, \boldsymbol{\theta}^{(t,n,\mathbf{x})}\}$ are large n samples successively generated by the Gibbs sampler from

$$P_{\boldsymbol{\Phi}^{(t)}}(\boldsymbol{\theta}|\mathbf{x}) = \frac{\exp \left\{ \beta^{(t)} \sum_{(s,r)} \theta_s \theta_r + \sum_{s \in S} h_s^{(t)} \theta_s \right\}}{Z \left(\beta^{(t)}, \{h_s^{(t)}\}_{s \in S} \right)},$$

with

$$\begin{aligned} h_s^{(t)} &= h^{(t)} - \log \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left\{ -\frac{(x_s - \mu_0)^2}{2\sigma_0^2} \right\} \right) \\ &+ \log \left(\sum_{l=1}^L \frac{p_l^{(t)}}{\sqrt{2\pi\sigma_l^{2(t)}}} \exp \left\{ -\frac{(x_s - \mu_l^{(t)})^2}{2\sigma_l^{2(t)}} \right\} \right) \end{aligned}$$

and $Z \left(\beta^{(t)}, \{h_s^{(t)}\}_{s \in S} \right)$ being the normalizing constant, and $\{\boldsymbol{\theta}^{(1,\boldsymbol{\varphi})}, \dots, \boldsymbol{\theta}^{(n,\boldsymbol{\varphi})}\}$ are generated from $P_{\boldsymbol{\varphi}}(\boldsymbol{\theta})$. Here for vector \mathbf{v} , $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^T$. Similarly,

$$\frac{C}{Z(\boldsymbol{\varphi})} = E_{\boldsymbol{\varphi}}[\exp\{-\boldsymbol{\varphi}^T \mathbf{H}(\boldsymbol{\Theta})\}] \approx \frac{1}{n} \sum_{i=1}^n \exp\{-\boldsymbol{\varphi}^T \mathbf{H}(\boldsymbol{\theta}^{(i,\boldsymbol{\varphi})})\},$$

where C is the number of all possible configurations $\boldsymbol{\theta}$ of $\boldsymbol{\Theta}$. Then the difference between Q_2 -functions in the Armijo condition can be approximated by

$$\begin{aligned} Q_2(\boldsymbol{\varphi}^{(t+1,m)}|\boldsymbol{\Phi}^{(t)}) - Q_2(\boldsymbol{\varphi}^{(t)}|\boldsymbol{\Phi}^{(t)}) \\ \approx \frac{1}{n} (\boldsymbol{\varphi}^{(t+1,m)} - \boldsymbol{\varphi}^{(t)})^T \sum_{i=1}^n \mathbf{H}(\boldsymbol{\theta}^{(t,i,\mathbf{x})}) \\ + \log \left(\frac{\sum_{i=1}^n \exp\{-\boldsymbol{\varphi}^{(t+1,m)T} \mathbf{H}(\boldsymbol{\theta}^{(t,i,\mathbf{x})})\}}{\sum_{i=1}^n \exp\{-\boldsymbol{\varphi}^{(t)T} \mathbf{H}(\boldsymbol{\theta}^{(t,i,\mathbf{x})})\}} \right). \end{aligned}$$

Back to $Q_1(\boldsymbol{\phi}|\boldsymbol{\Phi}^{(t)})$, the local conditional probability of $\boldsymbol{\Theta}$ given \mathbf{x} can also be approximated by the Gibbs sampler:

$$\gamma_s^{(t)}(i) = P_{\boldsymbol{\Phi}^{(t)}}(\Theta_s = i|\mathbf{x}) \approx \frac{1}{n} \sum_{k=1}^n \mathbf{1}(\boldsymbol{\theta}_s^{(t,k,\mathbf{x})} = i). \quad (12)$$

3.2. Implementation of the LIS-Based FDR Procedure

The algorithm for the LIS-based data-driven procedure, denoted as LIS for single group analysis, SLIS for separate analysis of multiple groups, and PLIS for pooled analysis for multiple groups, is given below:

1. Set initial values $\boldsymbol{\Phi}^{(0)} = \{\boldsymbol{\phi}^{(0)}, \boldsymbol{\varphi}^{(0)}\}$ for the model parameters $\boldsymbol{\Phi}$ of each group;
2. Update $\boldsymbol{\phi}^{(t)}$ from equations (6), (7) and (8);
3. Update $\boldsymbol{\varphi}^{(t)}$ from equations (10) and (11);
4. Iterate Steps 2 and 3 until convergence, then obtain the estimate $\hat{\boldsymbol{\Phi}}$ of $\boldsymbol{\Phi}$;
5. Plug-in $\hat{\boldsymbol{\Phi}}$ to obtain the test statistics \widehat{LIS} from equation (12);
6. Apply the data-driven procedure (LIS, SLIS, or PLIS).

The GEM algorithm is stopped when the following stopping rule

$$\max_i \left(\frac{|\Phi_i^{(t+1)} - \Phi_i^{(t)}|}{|\Phi_i^{(t)}| + \epsilon_1} \right) < \epsilon_2, \quad (13)$$

where Φ_i is the i th coordinate of vector $\boldsymbol{\Phi}$, is satisfied for three consecutive regular Newton-Raphson iterations with $m = 0$ in (10), or the prespecified maximum number of iterations is reached. Stopping rule (13) was applied by Booth and Hobert (1999) to the Monte Carlo EM method, where they set $\epsilon_1 = 0.001$, ϵ_2 between 0.002 and 0.005, and the rule to be satisfied for three consecutive iterations to avoid stopping the algorithm prematurely because of Monte Carlo error.

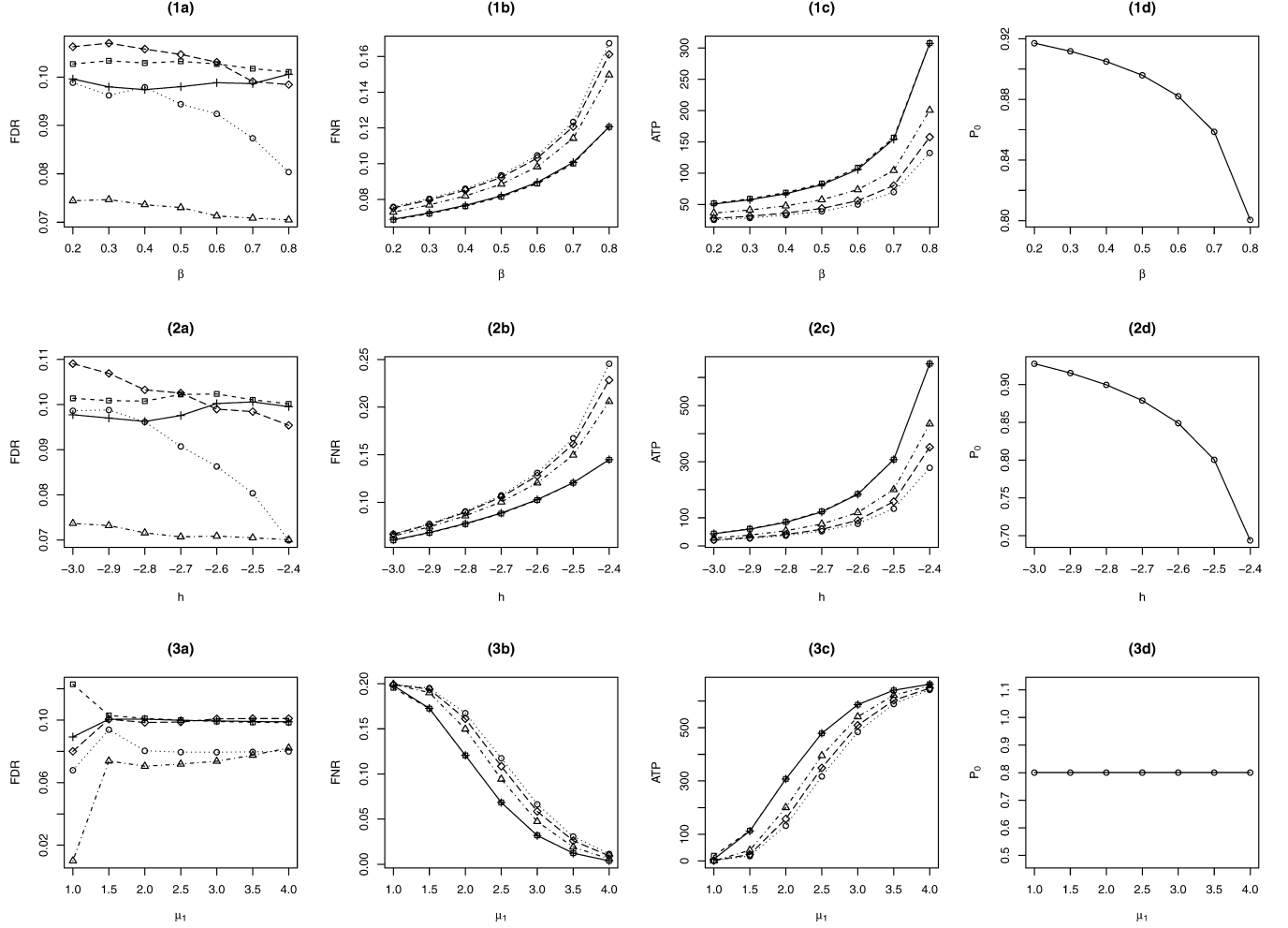


Figure 1. Comparison of BH (\circ), q -value (\diamond), Lfdr (\triangle), OR ($+$), and LIS (\square) for a single group with $L = 1$.

We used $\epsilon_1 = \epsilon_2 = 0.001$ in simulation studies and real-data analysis. Constant $\alpha = 10^{-4}$ is recommended by Nocedal and Wright (2006) for the Armijo condition (11), and the Newton–Raphson step length in (10) is halved by using $\lambda_m = 2^{-m}$. In practice, the Armijo condition (11) might not be satisfied when the step length $\|\varphi^{(t+1,m)} - \varphi^{(t)}\|$ is very small. In this situation, the iteration within Step 3 is stopped by an alternative criterion

$$\max_i \left(\frac{|\varphi_i^{(t+1,m)} - \varphi_i^{(t)}|}{|\varphi_i^{(t)}| + \epsilon_1} \right) < \epsilon_3$$

with $\epsilon_3 < \epsilon_2$, for example, $\epsilon_3 = 10^{-4}$ if $\epsilon_2 = 0.001$. Small a and b should be chosen in (8). We choose $a = 1$ and $b = 2$.

4. Simulation Studies

The simulation setups are similar to those in Sun and Cai (2009) and Wei et al. (2009), but with 3D data. The performances of the proposed LIS-based oracle (OR) and data-driven procedures are compared with the BH approach

(Benjamini and Hochberg, 1995), the q -value procedure (Storey, 2003), and the local FDR (Lfdr) procedure (Sun and Cai, 2007) for single group analysis; and the performances of SLIS and PLIS are compared with BH, q -value, and the conditional Lfdr (CLfdr) procedure (Cai and Sun, 2009) for multiple groups. The Lfdr and CLfdr procedures are shown to be optimal for independent tests (Sun and Cai, 2007; Cai and Sun, 2009). For simulations with multiple groups, all the procedures are globally implemented using all the locally computed test statistics based on each method from each group. The q -values are obtained using the R package `qvalue` (Dabney and Storey, 2014). For the Lfdr or CLfdr procedure, we use the proportion of the null cases generated from the Ising model with given parameters as the estimate of the probability of the null cases $P(\Theta_i = 0)$, together with the given null and nonnull distributions without estimating their parameters. For the LIS-based data-driven procedures, the maximum number of GEM iterations is set to be 1000 with $\epsilon_1 = \epsilon_2 = 0.001$, $\epsilon_3 = \alpha = 10^{-4}$, $a = 1$ and $b = 2$. For the Gibbs sampler, 5000 samples are generated from 5000 iterations after a burn-in period of 1000 iterations. In all simulations, each HMRf is on a $N = 15 \times 15 \times 15$ cubic lattice S , the

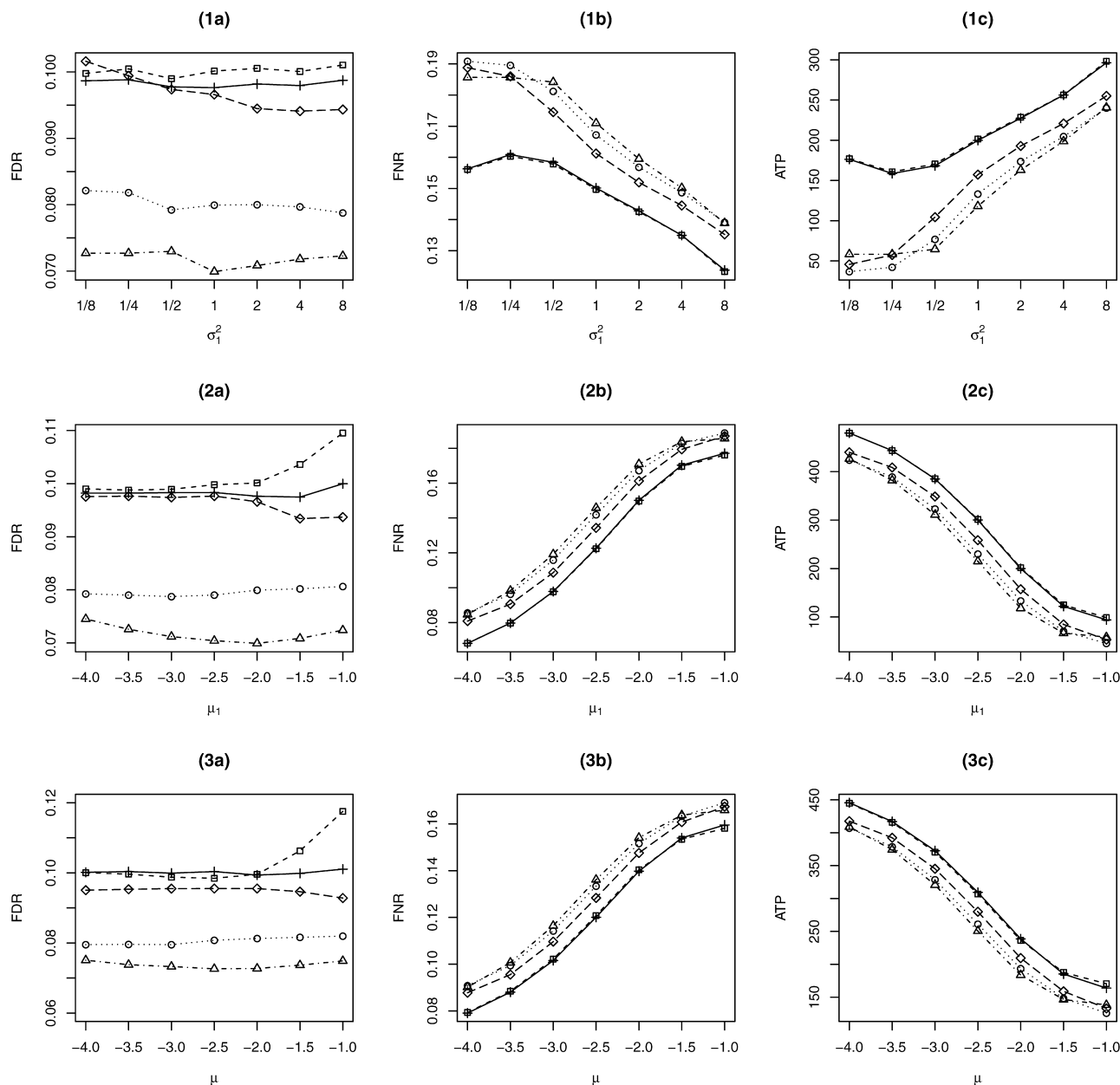


Figure 2. Comparison of BH (○), q -value (◇), Lfdr (△), OR (+), and LIS (□) for a single group with $L = 2$ (see 1a–2c), and the one with L being misspecified (see 3a–c).

number of replications $M = 200$ is the same as that in Wei et al. (2009), and the nominal FDR level is set at 0.10.

4.1. Single-Group Analysis

4.1.1. Study 1: $L = 1$. The MRF $\Theta = \{\Theta_s : s \in S\}$ is generated from the Ising model (1) with parameters (β, h) , and the observations $\mathbf{X} = \{X_s : s \in S\}$ are generated conditionally on Θ from $X_s | \Theta_s \sim (1 - \Theta_s)N(0, 1) + \Theta_s N(\mu_1, \sigma_1^2)$. Note that the MRF Θ is not observable in practice. Figure 1 shows the comparisons of the performance of BH, q -value, Lfdr, OR and LIS. In Figure 1(1a–1c), we fix $h = -2.5$, set $\mu_1 = 2$ and $\sigma_1^2 = 1$, and plot FDR, FNR, and the average number of true positives (ATP) yielded by these procedures as functions of

β . In Figure 1(2a–2c), we fix $\beta = 0.8$, set $\mu_1 = 2$ and $\sigma_1^2 = 1$, and plot FDR, FNR and ATP as functions of h . In Figure 1(3a–c), we fix $\beta = 0.8$ and $h = -2.5$, set $\sigma_1^2 = 1$, and plot FDR, FNR and ATP as functions of μ_1 . The corresponding average proportions of the nulls, denoted by P_0 , for each Ising model are given in Figure 1(1d–3d). The initial values for the numerical algorithm are set at $\beta^{(0)} = h^{(0)} = 0$, $\mu_1^{(0)} = \mu_1 + 1$ and $\sigma_1^{2(0)} = 2$.

From Figure 1(1a–3a), we can see that the FDR levels of all five procedures are controlled around 0.10 except one case of the LIS procedure in Figure 1(3a) with the lowest μ_1 , whereas the BH and Lfdr procedures are generally conservative. This case of obvious deviation of the LIS procedure is likely caused

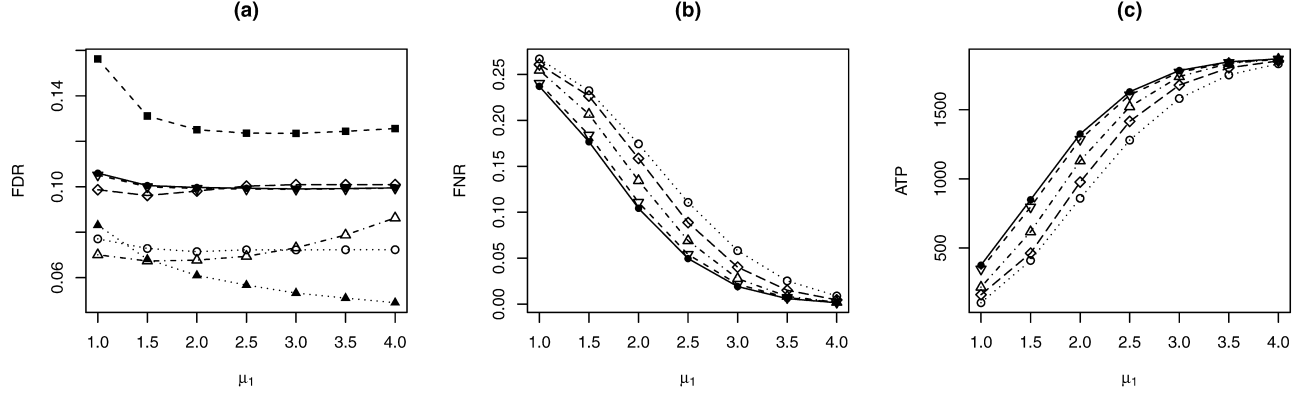


Figure 3. Comparison of BH (\circ), q -value (\diamond), CLfdr (\triangle), SLIS (∇), and PLIS (\bullet) for two groups with $L = 1$. In (a), \blacksquare and \blacktriangle represent the results by PLIS for each individual group; for PLIS, while the global FDR is controlled, individual-group FDRs may vary.

by the small lattice size N . As a confirmation, additional simulations by increasing the lattice size N to $30 \times 30 \times 30$ yield an FDR of 0.1019 for the same setup. From Figure 1(1b–3b) and (1c–3c) we can see that the two curves of OR and LIS procedures are almost identical, indicating that the data-driven LIS procedure works equally well as the OR procedure. These plots also show that the LIS procedure outperforms BH, q -value and Lfdr procedures with increased margin of performance in FNR and ATP as β or h increases or μ_1 is at a moderate level. Note that from Web Appendix A, we can see that β controls how likely the same-state cases cluster together, and (β, h) together control the proportion of the aggregation of nonnulls relative to that of nulls.

4.1.2. Study 2: $L = 2$. We now consider the case where the nonnull distribution is a mixture of two normal distributions. The MRF is generated from the Ising model (1) with fixed parameters $\beta = 0.8$ and $h = -2.5$, and the nonnull distribution is a two-component normal mixture $p_1 N(\mu_1, \sigma_1^2) + p_2 N(\mu_2, \sigma_2^2)$ with fixed $p_1 = p_2 = 0.5$, $\mu_2 = 2$, and $\sigma_2^2 = 1$. In Figure 2(1a–c), σ_1^2 varies from 0.125 to 8, and $\mu_1 = -2$. In Figure 2(2a–c), we fix $\sigma_1^2 = 1$ and vary μ_1 from -4 to -1 . The initial values are set at $\beta^{(0)} = h^{(0)} = 0$, $p_1^{(0)} = 1 - p_2^{(0)} = 0.3$, $\mu_1^{(0)} = \mu_1 + 1$, and $\sigma_1^{2(0)} = \sigma_1^2 + 1$, $l = 1, 2$.

Similar to Figure 1, we can see that the FDR levels of all the procedures are controlled around 0.10, where BH and Lfdr are conservative, and OR and LIS perform similarly and outperform the other three procedures. In Figure 2(2a) at $\mu_1 = -1$, additional simulations yield an FDR of 0.1035 when the lattice size N is increased to $30 \times 30 \times 30$ for the same setup.

The results from both simulation studies are very similar to those in Sun and Cai (2009) for the one-dimensional case using HMC. It is clearly seen that, for dependent tests, incorporating dependence structure into a multiple-testing procedure improves efficiency dramatically.

4.1.3. Study 3: misspecified nonnull. Following Sun and Cai (2009), we consider the true nonnull distribution to be the three-component normal mixture $0.4N(\mu, 1) + 0.3N(1, 1) + 0.3N(3, 1)$, but use a misspecified two component normal mixture $p_1 N(\mu_1, \sigma_1^2) + p_2 N(\mu_2, \sigma_2^2)$ in the LIS procedure. The

unobservable states are generated from the Ising model (1) with fixed parameters $\beta = 0.8$ and $h = -2.5$. The simulation results are displayed in Figure 2(3a–c), the true μ varies from -4 to -1 with increments of size 0.5. The initial values are set at $\beta^{(0)} = h^{(0)} = 0$, $p_1^{(0)} = p_2^{(0)} = 0.5$, $\mu_1^{(0)} = -\mu_2^{(0)} = -2$, and $\sigma_1^{2(0)} = 2$, $l = 1, 2$.

Figure 2(3a–c) shows that the LIS procedure performs similarly to OR under misspecified model. Additionally, the obvious biased FDR level by the LIS procedure at $\mu = -1$ reduces to 0.1067 when the lattice size N is increased to $30 \times 30 \times 30$.

4.2. Multiple-Group Analysis

Voxels in a human brain can be naturally grouped into multiple functional regions. For simulations with grouped multiple tests, we consider two lattice groups each with size $15 \times 15 \times 15$. The corresponding MRFs $\Theta_1 = \{\Theta_{1s} : s \in S\}$ and $\Theta_2 = \{\Theta_{2s} : s \in S\}$ are generated from the Ising model (1) with parameters $(\beta_1 = 0.2, h_1 = -1)$ and $(\beta_2 = 0.8, h_2 = -2.5)$, respectively. The observations $X_k = \{X_{ks}, s \in S\}$ are generated conditionally on Θ_k , $k = 1, 2$, from $X_{ks} | \Theta_{ks} \sim (1 - \Theta_{ks})N(0, 1) + \Theta_{ks}N(\mu_k, \sigma_k^2)$, where μ_1 varies from 1 to 4 with increments of size 0.5, $\mu_2 = \mu_1 + 1$ and $\sigma_1^2 = \sigma_2^2 = 1$. The initial values are $\beta_1^{(0)} = \beta_2^{(0)} = h_1^{(0)} = h_2^{(0)} = 0$, $\mu_2^{(0)} = \mu_1^{(0)} = \mu_1 + 1$, and $\sigma_1^{2(0)} = \sigma_2^{2(0)} = 2$.

The simulation results are presented in Figure 3, which are similar to that in Wei et al. (2009) for the one-dimensional case with multiple groups using HMCs. Figure 3(a) shows that all procedures are valid in controlling FDR at the pre-specified level of 0.10, whereas BH and CLfdr procedures are conservative. We also plot the within-group FDR levels of PLIS for each group separately. One can see that in order to minimize the global FNR level, the PLIS procedure may automatically adjust the FDRs of each individual group, either inflated or deflated reflecting the group heterogeneity, while the global FDR is appropriately controlled. In Figure 3(b) and (c) we can see that both SLIS and PLIS outperform BH, q -value and CLfdr procedures, indicating that utilizing the dependency information can improve the efficiency of a testing procedure, and the improvement is more evident for weaker signals (smaller values of μ_1). Between the two LIS-

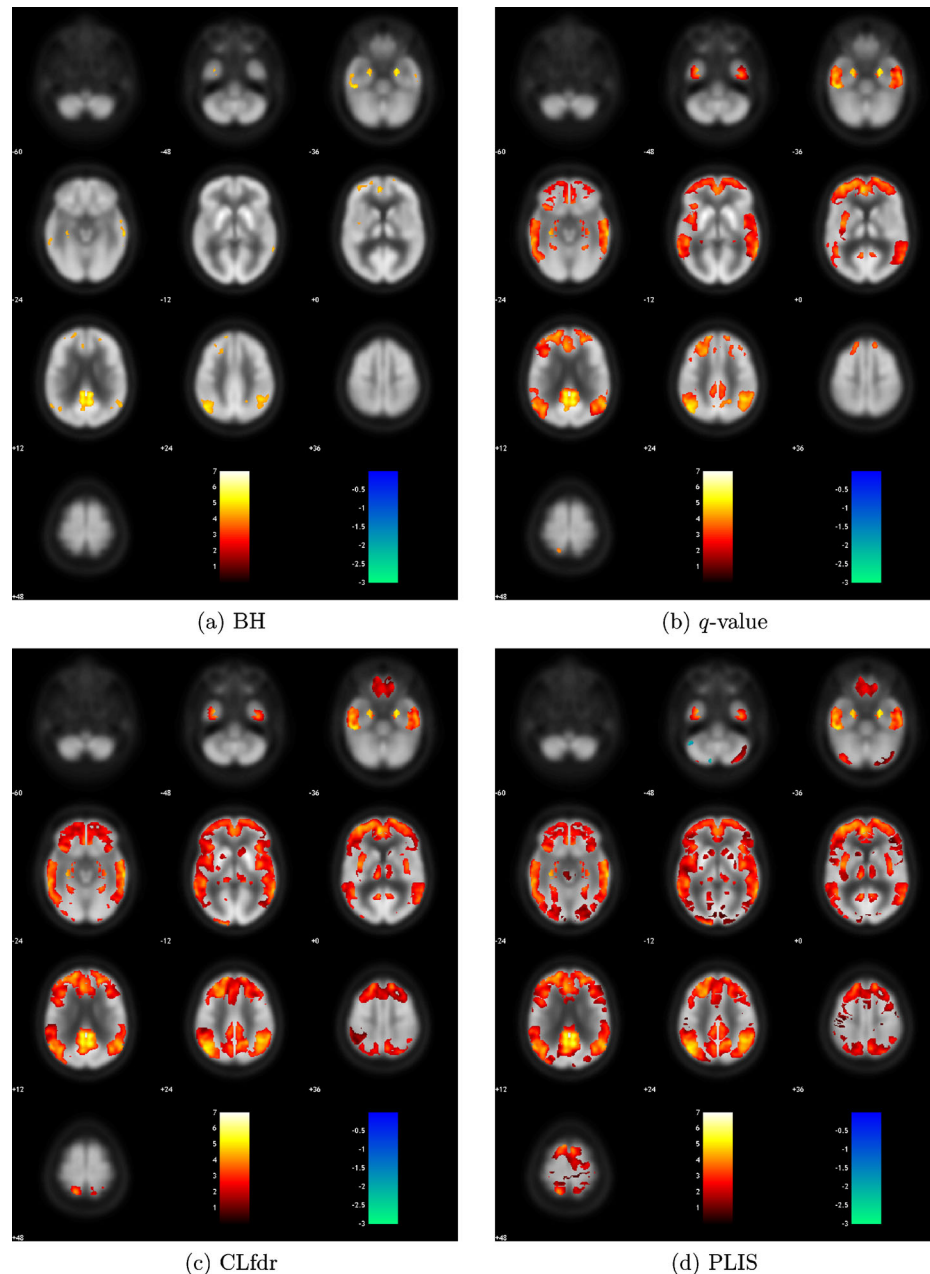


Figure 4. Z-values of the signals found by each procedure for the comparison between NC and MCI.

based procedures, PLIS slightly outperforms SLIS, indicating the benefit of ranking the LIS test statistics globally. In particular, ATP is 8.3% higher for PLIS than for SLIS when $\mu_1 = 1$.

5. ADNI FDG-PET Image Data Analysis

Alzheimer's disease (AD) is the most common cause of dementia in the elderly population. Much progress has been made in the diagnosis of AD including clinical assessment and neuroimaging techniques. One such extensively used neuroimaging technique is FDG-PET imaging, which is used to evaluate the cerebral metabolic rate of glucose (CMRgl). We consider the FDG-PET image data from the ADNI database (adni.loni.usc.edu) as an illustrative example.

The data set consists of the baseline FDG-PET images of 102 normal control (NC) subjects and 206 patients with mild cognitive impairment (MCI), a prodromal stage of AD. Sixty one brain regions of interest (ROIs) are considered (see Web Appendix E for details), where the number of voxels in each region ranges from 149 to 20,680 with a median of 2,517. The total number of voxels of these 61 ROIs is $N = 251,500$. The goal is to identify voxels with reduced CMRgl in MCI patients comparing to NC.

We apply the HMRF-PLIS procedure to the ADNI data, and compare to BH, q -value and CLfdr procedures. We implement the BH procedure globally for the 61 ROIs, whereas we treat each region as a group for the q -value, CLfdr and

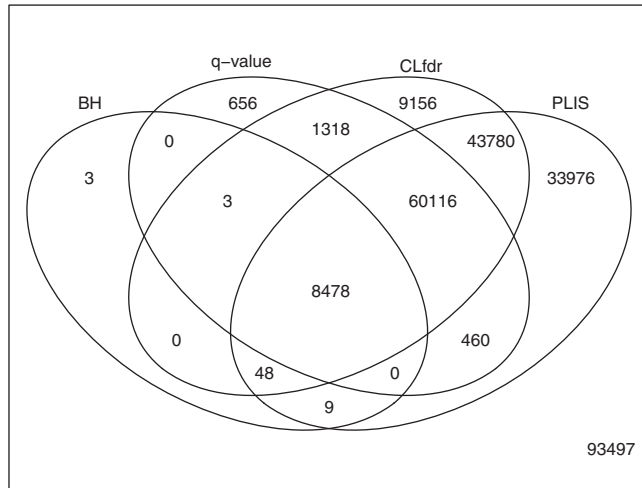


Figure 5. Venn diagram for the number of signals found by each procedure for the comparison between NC and MCI. Number of signals discovered by each procedure: BH=8,541, q -value=71,031, CLfdr=122,899, and PLIS=146,867.

PLIS procedures. For the BH and q -value procedures, a total number of N two-sample Welch's t -tests (Welch, 1947) are performed, and their corresponding two-sided p -values are obtained. For the PLIS and CLfdr procedures, z -values are used as the observed data \mathbf{x} , which are obtained from those t statistics by the transformation $z_i = \Phi^{-1}[G_0(t_i)]$, where Φ and G_0 are the cumulative distribution functions of the standard normal and the t statistic, respectively. The null distribution is assumed to be the standard normal distribution. The nonnull distribution is assumed to be a two-component normal mixture for PLIS. The LIS statistics in the PLIS procedure are approximated by 10^6 Gibbs-sampler samples, and the Lfdr statistics in the CLfdr procedure are computed by using the R code of Sun and Cai (2007). All the four testing procedures are controlled at a nominal FDR level of 0.001. In the GEM algorithm for HMRF estimation, the initial values for β and h in the Ising model are set to be zero. The initial values for the nonnull distributions are estimated from the signals claimed by BH at an FDR level of 0.1. The maximum number of GEM iterations is set to be 5000 with $\epsilon_1 = \epsilon_2 = 0.001$, $\epsilon_3 = \alpha = 10^{-4}$, $a = 1$, and $b = 2$. For the Gibbs sampler embedded in the GEM, 5000 samples are generated from 5000 iterations after a burn-in period of 1000 iterations. In this data analysis, the GEM algorithm reaches the maximum iteration and is then claimed to be converged for five ROIs. Among all 61 ROIs, the estimates of β have a median of 1.57 with the interquartile range of 0.36, and the estimates of h have a median of -3.71 with the interquartile range of 1.52. Such magnitude of parameter variation supports the multi-region analysis of the ADNI FDG-PET image data because even a 0.1 difference in β or h can result in quite different Ising models, see Figure 1(d) and (2d).

Figure 4 shows the z -values (obtained by comparing CM-Rgl values between NC and MCI) of all the signals claimed by each procedure. Figure 5 summarizes the number of voxels that are claimed as signals by each procedure. We can see that PLIS finds the largest number of signals and cov-

ers 91.5%, 97.2%, and 99.9% of signals detected by CLfdr, q -value, and BH, respectively. It is interesting to see that the PLIS procedure finds more than 17 times signals as BH, twice as many signals as q -value, and about 20% more signals than the CLfdr procedure.

Detailed interpretations of the scientific findings are provided in Web Appendix E.

6. Concluding Remarks

In this article, we consider LIS-based FDR procedures based on HMRF for 3D neuroimage data, where HMRF provides a natural way of modeling spatial correlations. The procedures aim to minimize the FNR while FDR is controlled at a pre-specified level. We find brain regions are spatially heterogeneous, hence model each region separately by a single HMRF, and implement the PLIS procedure to minimize the global FNR. We propose a GEM algorithm based on the penalized likelihood to obtain the HMRF parameter estimates, which overcomes the unboundedness of the original likelihood function. Numerical analysis shows the superiority of the HMRF-LIS-based procedures over commonly used FDR procedures, illustrating the value of HMRF-LIS-based FDR procedures for spatially correlated image data. The asymptotic properties of the PMLE of HMRF and the data-driven HMRF-LIS-based procedures are of interest for future research.

7. Supplementary Materials

Web Appendix A mentioned in Sections 2 and 4, Web Appendices B-D referenced in Section 3, Web Appendix E mentioned in Section 5, and a MATLAB package implementing the proposed FDR procedure are available with this paper at the *Biometrics* website on Wiley Online Library.

ACKNOWLEDGEMENTS

We are grateful to Dr. Jeanine Houwing-Duistermaat, an Associate Editor and two anonymous referees for their helpful comments. The research is supported in part by NIH grant R01-AG036802 and NSF grants DMS-1007590 and DMS-1407142.

We also would like to thank ADNI for providing the brain image data that were obtained from the ADNI database (adni.loni.usc.edu). Data collection and sharing was funded by NIH grant U01-AG024904 and DOD grant W81XWH-12-2-0012. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; ; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health

Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**, 289–300.
- Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25**, 60–83.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188.
- Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265–285.
- Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. New York: Springer.
- Cai, T. and Sun, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *Journal of the American Statistical Association* **104**, 1467–1481.
- Chandgotia, N., Han, G., Marcus, B., Meyerovitch, T., and Pavlov, R. (2014). One-dimensional Markov random fields, Markov chains and topological Markov fields. *Proceedings of the American Mathematical Society* **142**, 227–242.
- Chen, J., Tan, X., and Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statistica Sinica* **18**, 443–465.
- Chumbley, J. R. and Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage* **44**, 62–70.
- Chumbley, J., Worsley, K., Flandin, G., and Friston, K. (2010). Topological FDR for neuroimaging. *NeuroImage* **49**, 3057–3064.
- Ciuperca, G., Ridolfi, A., and Idier, J. (2003). Penalized maximum likelihood estimator for normal mixtures. *Scandinavian Journal of Statistics* **30**, 45–59.
- Dabney, A. and Storey, J. D. (2014). qvalue: Q-value estimation for false discovery rate control. *R package version 1.36.0*.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association* **99**, 96–104.
- Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics* **34**, 275–297.
- Garey, L. J. (2006). *Brodmann's Localisation in the Cerebral Cortex*. New York: Springer.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**, 870–878.
- Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society, Series B* **64**, 499–517.
- Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *The Annals of Statistics* **32**, 1035–1061.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. New York: Springer.
- Huang, L., Goldsmith, J., Reiss, P. T., Reich, D. S., and Crainiceanu, C. M. (2013). Bayesian scalar-on-image regression with application to association between intracranial DTI and cognitive outcomes. *NeuroImage* **83**, 210–223.
- Johnson, T. D., Liu, Z., Bartsch, A. J., and Nichols, T. E. (2013). A Bayesian non-parametric Potts model with application to pre-surgical fMRI data. *Statistical Methods in Medical Research* **22**, 364–381.
- Magder, L. S. and Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association* **91**, 1141–1151.
- Nocedal, J. and Wright, S. (2006). *Numerical Optimization*, 2nd edition. New York: Springer.
- Ridolfi, A. (1997). Maximum likelihood estimation of hidden Markov model parameters, with application to medical image segmentation. Tesi di Laurea, Politecnico di Milano, Milan, Italy.
- Roberts, G. O. and Smith, A. F. M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* **49**, 207–216.
- Stoer, J. and Bulirsch, R. (2002). *Introduction to Numerical Analysis*, 3rd edition. New York: Springer.
- Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q -value. *The Annals of Statistics* **31**, 2013–2035.
- Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association* **102**, 901–912.
- Sun, W. and Cai, T. T. (2009). Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society, Series B* **71**, 393–424.
- Wei, Z., Sun, W., Wang, K., and Hakonarson, H. (2009). Multiple testing in genome-wide association studies via hidden Markov models. *Bioinformatics* **25**, 2802–2808.
- Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* **34**, 28–35.
- Winkler, G. (2003). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd edition. New York: Springer.
- Wu, W. B. (2008). On false discovery control under dependence. *The Annals of Statistics* **36**, 364–380.
- Zhang, C., Fan, J., and Yu, T. (2011). Multiple testing via FDR.L for large-scale image data. *The Annals of Statistics* **39**, 613–642.
- Zhang, X., Johnson, T. D., Little, R. J. A., and Cao, Y. (2008). Quantitative magnetic resonance image analysis via the EM algorithm with stochastic variation. *The Annals of Applied Statistics* **2**, 736–755.

Received June 2014. Revised March 2015. Accepted April 2015.